

データベース

情報理工学院 情報工学系

横田 治夫

西8E 705号室

yokota@cs.titech.ac.jp

本日(5月4日)の内容

- 接続状況確認
- 講義のイントロダクション
 - データベースに対するイメージをもってもら
 - 世の中の現状を鑑みて
 - 本講義の位置づけ
 - 取り扱う内容と重要な概念
 - 講義資料とOCWについて
 - 毎年それなりに変更を加えている
 - 次回からは自分でプリントして持ってくるように
 - 講義日程と成績評価について
- データモデルについて

2020/5/4

データベース (©横田)

2

身近なデータベース

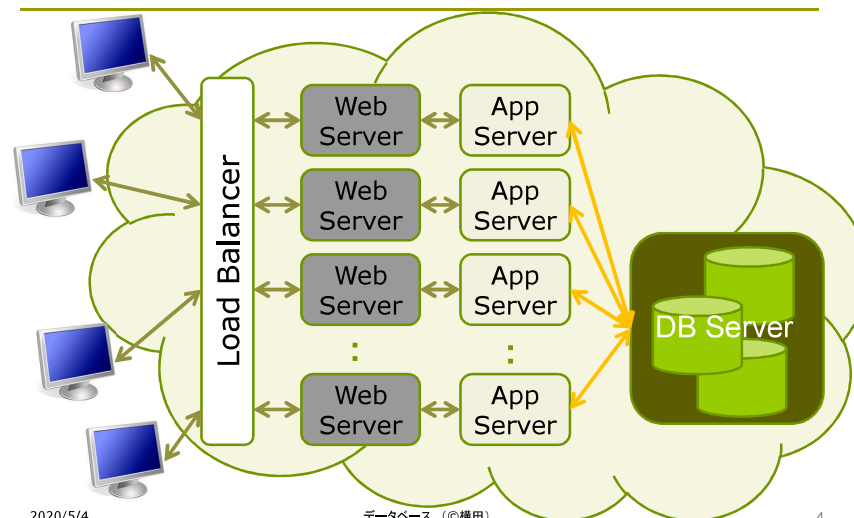
- 銀行の口座データベース
 - キャッシュディスペンサー、ATM 経由でアクセス
- 学校の学生データベース(成績管理)
- 企業の人事データベース(給与・勤怠管理)
- 工場の機材データベース(在庫管理)
- 自治体の住民情報データベース(住民票発行)
- 図書館の文献データベース
- スーパーマーケットの商品データベース
- ...
 - 各種サービスのWebの裏にはあると思って間違いない

2020/5/4

データベース (©横田)

3

Web の裏には DB あり



2020/5/4

データベース (©横田)

4

ちょっと考えてみよう(1-1)

- 自分の身の回りで、例に出てこなかったデータベースは何か思いつきますか？
 - ○○○に使われているデータベース
 - 何でもよいので、思いついたものを1つフォームに書き込みましょう

データベースの大きさのイメージ

- データは増え続けている！ (cf. ビッグデータ)
- 遺伝子データベース(DNA 配列): 2 TB (2×10^{12} B)
- 米国国会図書館: 20 TB (2×10^{13} B)
- 衛星からの観測データベース: 年間 6 PB (6×10^{15} B)
- スーパーマーケットの履歴データベース
 - POS (Point of Sales) による在庫管理 → 戦略決定
 - データウェアハウス(Data Warehouse)と呼ばれる
 - ログに対するデータマイニング、OLAP (On Line Analytical Proc.)
 - 1件100B のデータのトランザクションを1秒間に1,000件
 - 1秒間に 100KB、1時間で 360 MB、1日 10GB、1年間 3TB
 - 米国 Walmart のデータウェアハウス
 - 1992:1TB, 1997:24TB, 1999:101TB, 2004:570TB, 2008:2.5PB

単位(参考)

□ 10^1	Deka (da)	□ 10^{-1}	Deci (d)
□ 10^2	Hecto (h)	□ 10^{-2}	Centi (c)
□ 10^3	Kilo (k)	□ 10^{-3}	Milli (m)
□ 10^6	Mega (M)	□ 10^{-6}	Micro (μ)
□ 10^9	Giga (G)	□ 10^{-9}	Nano (n)
□ 10^{12}	Tera (T)	□ 10^{-12}	Pico (p)
□ 10^{15}	Peta (P)	□ 10^{-15}	Femto (f)
□ 10^{18}	Exa (E)	□ 10^{-18}	Atto (a)
□ 10^{21}	Zetta (Z)	□ 10^{-21}	Zepto (z)
□ 10^{24}	Yotta (Y)	□ 10^{-24}	Yocto (y)

別な例: facebook

- 非常に大量のデータが日々生成・蓄積
 - 既にユーザー数は24億以上、年間成長率 8%
 - 約16億の利用者は毎日アップデート
 - 毎月10億を超える写真がアップロード
 - 毎月1千万を超えるビデオがアップロード
 - Web リンク、ブログポスト、メモ等大量のデータ
 - 毎日48億件のコンテンツ
 - 毎日600TB以上データを処理。
 - トータルで300PBを処理
- 彼らは巨大なデータセンターを持つ

データベースシステムに要求される機能

ー 集められ、整理されたデータは大きな財産 ー

- 使いやすく、高速に検索・更新するための機能
- 異なる利用形態に対応する機能
 - 様々なアプリケーションから利用
- データを共有するための機能
- 故障や異常が生じた場合にも内容を失わないための機能
- より高度な機能

ファイルとデータベースの違い

2020/5/4

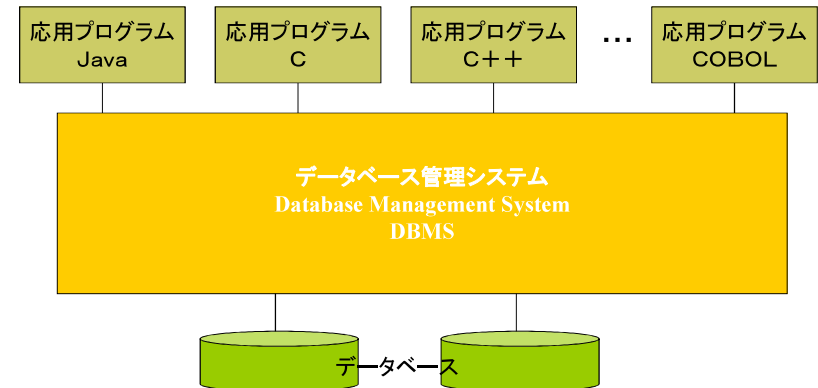
データベース (©横田)

9

データベース管理システム

DBMS: Database Management System

- 前述の機能を実現するためのシステム



2020/5/4

データベース (©横田)

10

DBMSの機能を大きく分けると

- データの操作
 - 体系的な操作を可能に
 - データモデル、問い合わせ言語
 - データを構造化して格納・検索
 - データの格納方法、インデックス等
 - 20TB を 20MB/s で単に読むだけで
- トランザクションの管理
 - 複数のデータ操作が着実に行われるように管理
 - 同じデータベースを同時に利用できるように
 - 故障や異常が発生しても大丈夫なように管理

2020/5/4

データベース (©横田)

11

ちょっと考えてみよう(1-2)

- データを単にファイルとして格納した場合
 - 形式は、テキストでもエクセルのようなCSVでもOK
 - 何が困るか、困ることを3つ挙げてみましょう

2020/5/4

データベース (©横田)

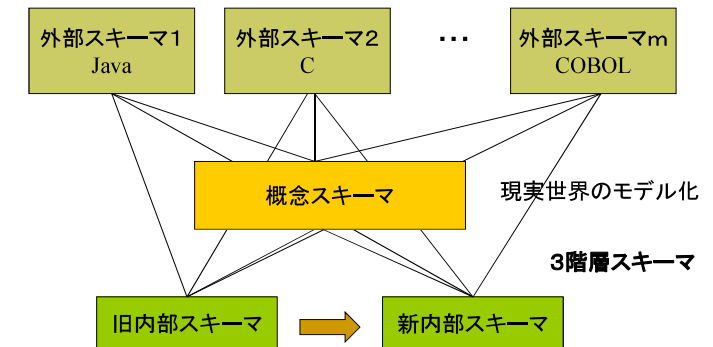
12

データモデル

- 現実世界をシステム内で表現するための記号系
 - 銀行に預けたお金 → 名前とID と金額
 - 学生の成績 → 名前と学籍番号と点数
 - スーパーマーケット → 商品名と価格と在庫
- それをシステムの中でどう表現すればよいのか
 - 探したい情報を適切に手に入りたい
 - 様々なアプリケーションから使いたい
 - 安定して使い続けたい(システムに変更があっても)

スキーマ階層とデータモデル

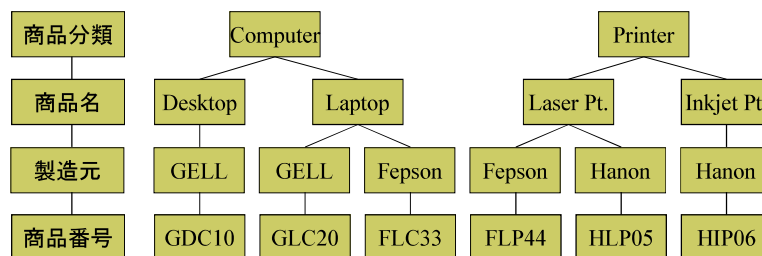
- 概念階層の導入とデータ独立



- 応用プログラムをデータベースの変更から安定に保つ

データベースのモデル1

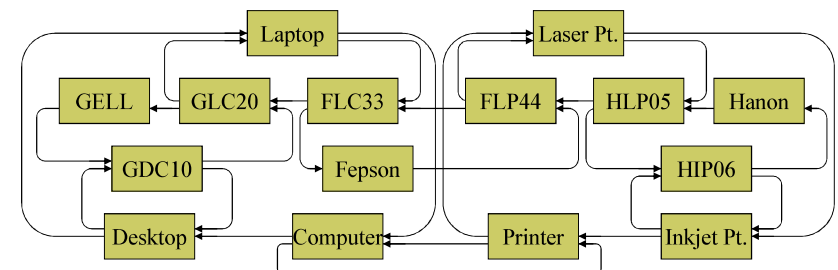
- 階層型モデル



- 人間の直感に近い
- 決められたパターン以外の検索に弱い

データベースのモデル2

- ネットワーク型モデル



- 柔軟な検索パターンに対応可能
- サイズが大きくなると複雑で全体の理解が困難

データベースのモデル3

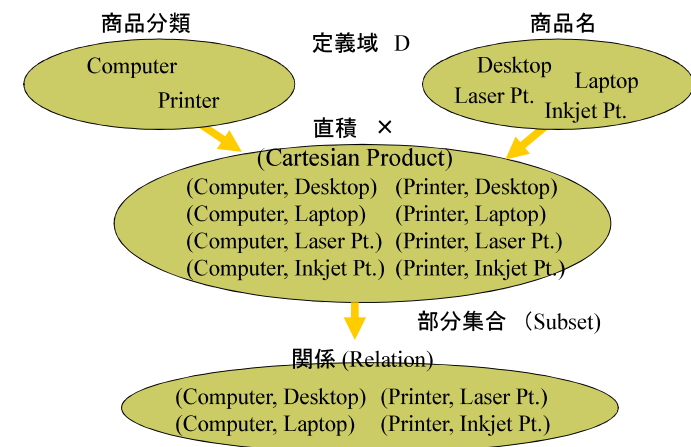
- 関係データモデル
- 1970年に E. F. Codd (当時 IBM) が提案
 - 2003年4月18日没
 - 1923年生まれ、享年79歳
- 集合論をベース
- 定義: $t \in R \subset D_1 \times D_2 \times \dots \times D_n$
 - t: タプル Tuple (組)
 - R: 関係 Relation
 - D_i : 定義域 Domain
 - n: 次数 Degree

2020/5/4

データベース (©横田)

17

関係データモデル



2020/5/4

データベース (©横田)

18

関係の例1

製造元関係			スキーマ Schema
商品番号	商品名	製造元	属性名 Attribute Name
GDC10	Desktop	GELL	タプル Tuple
GLC20	Laptop	GELL	
FLC33	Laptop	Fepson	
FLP44	Laser Pt.	Fepson	
HLP05	Laser Pt.	Hanon	
HIP06	Inkjet Pt.	Hanon	
TLCD70	LCDisplay	Tenao	
			属性 Attribute
			Cardinality:7

2020/5/4

データベース (©横田)

19

関係の例2

商品分類関係		秋葉原支店関係		
商品分類	商品名	商品番号	販売価格	在庫
Computer	Desktop	GDC10	150,000	5
Computer	Laptop	GLC20	90,000	12
Printer	Laser Pt.	FLC33	70,000	3
Printer	Inkjet Pt.	HIP06	30,000	4
LCDisplay	600dpi LPR	TLCD70	40,000	2

2020/5/4

データベース (©横田)

20

関係データモデルの用語

- 見え方: 親しみやすいテーブル
- テーブル全体: 関係(Relation)
- 行: タプル(Tuple) / 列: 属性(Attribute)
 - 属性には名前が付く: 属性名(Attribute Name)
 - 属性名の並び: スキーマ(Schema)
 - タプル数: 濃度(Cardinality)
- 普通のテーブルとの違い
 - タプルの重複は許さない(許す場合: マルチセット)
 - タプル、属性の順番は無意味

キー(Key)

- 関係の中でタプルを一意に識別できる属性の集合
 - 冗長でない
 - 例:
 - 製造元関係の商品番号
 - 商品名と製造元の組み合わせ
- 複数の可能性: キー候補(Candidate Key)
 - 主キー／基本キー(Primary Key)と
 - 代替キー(Alternate Key)

null 値とキー制約

- null値(空値)
 - 属性が値を取り得ない
 - 属性は値を取り得るが、値が不明
 - そもそも値を取るかどうか分からない
- キー制約
 - 主キーの属性値は
 - 唯一で、null 値であってはならない

ちょっと考えてみよう(1-3)

- 他のモデルと比較して関係モデルのメリットは何？

トランザクション管理の役割(1)

- データを共有するための機能
 - 同時に複数の利用者にサービスを提供する
 - 同じデータを同時に利用した時の一貫性の保持
- 異なる利用形態に対応する機能
 - 複数のアプリケーションからの利用
 - 例: スーパーのPOS と戦略決定

トランザクション管理の役割(2)

- 故障や異常が生じた場合にも内容を失わないための機能
 - システム全体やその一部が故障してもデータを復旧
- 更新により矛盾を生じさせないための機能
 - 相互に関係しあうデータの一方の変更による不具合の解消
- 不正なアクセス対策やプライバシー保護のための機能
 - データベースへのアクセス権の管理

ちょっと考えてみよう(1-4)

- 再び単なるファイルとの違い
- エクセルでデータを管理したとして、トランザクション管理の中でできそうなことは何？

本講義の内容

- データの操作とトランザクション管理をメイン
 - データベース管理システム
 - データモデル
- 時間があったら新しい話題も
- 成績評価
 - 講義中の回答
 - 可能になったら期末試験を行う

具体的な講義の項目

- イン트로ダクション(本日)
- データベースのモデル(本日)
- 関係データベースの操作(関係代数、関係論理)
- 正規化
- 問い合わせ言語
- トランザクション処理
- データベースの内部構造
- データベースの問い合わせ処理
- その他(オブジェクト指向DB、XML/RDF-DB等々)

2020/5/4

データベース (©横田)

29

講義の予定(1)

- 5月 4日(月) 第1回(本日)
- 5月 7日(木) 第2回
- 5月11日(月) 第3回
- 5月14日(木) 第4回
- 5月18日(月) 第5回
- 5月21日(木) 第6回
- 5月25日(月) 第7回
- 5月28日(木) 第8回

2020/5/4

データベース (©横田)

30

講義の予定(2)

- 6月 1日(月) 第9回
- 6月 4日(木) 第10回
- 6月 8日(月) 第11回
- 6月11日(木) 第12回
- 6月15日(月) 第13回
- 6月18日(木) 第14回(期末試験)

2020/5/4

データベース (©横田)

31

参考文献(データ操作・モデルがメイン1)

- C. J. Date, "An Introduction to Database Systems", Addison-Wesley, 8th edition, 2007, ISBN:0-321-19784-4, 監訳: 藤原譲、データベースシステム概論(第6版の訳本)、丸善、ISBN:4-621-04276-9, 18,000円、英初版は1975年
- J. D. Ullman, "Principles of Database and Knowledge-Base Systems", Computer Science Press, 1989, Vol 1,2, ISBN:0-7167-8158-1 (J. D. Ullman, "Principles of Database Systems", Computer Science Press, 2nd edition, 1982, ISBN:0-914894-36-6, 邦訳: 国井利泰、大保信夫: データベースシステムの原理、日本コンピュータ協会)

2020/5/4

データベース (©横田)

32

参考文献(データ操作・モデルがメイン2)

- J. D. Ullman, J. Widom, “A First Course In Database Systems”, 3rd edition, Prentice Hall, 2007, ISBN: 01360063X, \$126.00
- 北川 博之, 「データベースシステム(改訂2版)」, オーム社, 2020, ISBN: 978-4274225161, 3,520円
- 植村 俊亮, 「入門データベース」, オーム社, 2018, ISBN: 978-4274222924, 2,420円
- 増永 良文, 「リレーショナルデータベース入門(第3版)」, サイエンス社, 2016, ISBN: 978-4-7819-1390, 3,200円
- 川越 恭二, 「楽しく学べるデータベース」, 共立出版, 2014, ISBN: 978-4-320-12378-6, 2,400円

参考文献(トランザクション)

- J. Gray, A. Reuter, “Transaction Processing: Concepts and Techniques,” Morgan Kaufmann, 1993, ISBN: 1-55860-190-2, 監訳: 喜連川優、トランザクション処理—概念と技法—(上下巻)、日経BP、2001、上下巻とも6,800 円
- B. Claybrook, “OLTP: Online Transaction Processing Systems”, John Wiley, 1992, ISBN: 0-471-55668-8, 監訳: 信国弘毅、OLTPオンライントランザクション処理システム、富士ソフトウェア、5,150円

本日の講義の最後に

- 初めてのオンライン講義でこなれていませんが、コメントがあったら書き込んで下さい(オプションル)